



ML-SIM: universal reconstruction of structured illumination microscopy images using transfer learning

CHARLES N. CHRISTENSEN,^{1,2}  EDWARD N. WARD,¹ MENG LU,¹ PIETRO LIO,² AND CLEMENS F. KAMINSKI^{1,*}

¹University of Cambridge, Department of Chemical Engineering and Biotechnology, Laser Analytics Group, Philippa Fawcett Dr, Cambridge, UK

²University of Cambridge, Department of Computer Science and Technology, Artificial Intelligence Group, JJ Thomson Ave, Cambridge, UK

*Corresponding author: cfk23@cam.ac.uk

Abstract: Structured illumination microscopy (SIM) has become an important technique for optical super-resolution imaging because it allows a doubling of image resolution at speeds compatible with live-cell imaging. However, the reconstruction of SIM images is often slow, prone to artefacts, and requires multiple parameter adjustments to reflect different hardware or experimental conditions. Here, we introduce a versatile reconstruction method, ML-SIM, which makes use of transfer learning to obtain a parameter-free model that generalises beyond the task of reconstructing data recorded by a specific imaging system for a specific sample type. We demonstrate the generality of the model and the high quality of the obtained reconstructions by application of ML-SIM on raw data obtained for multiple sample types acquired on distinct SIM microscopes. ML-SIM is an end-to-end deep residual neural network that is trained on an auxiliary domain consisting of simulated images, but is transferable to the target task of reconstructing experimental SIM images. By generating the training data to reflect challenging imaging conditions encountered in real systems, ML-SIM becomes robust to noise and irregularities in the illumination patterns of the raw SIM input frames. Since ML-SIM does not require the acquisition of experimental training data, the method can be efficiently adapted to any specific experimental SIM implementation. We compare the reconstruction quality enabled by ML-SIM with current state-of-the-art SIM reconstruction methods and demonstrate advantages in terms of generality and robustness to noise for both simulated and experimental inputs, thus making ML-SIM a useful alternative to traditional methods for challenging imaging conditions. Additionally, reconstruction of a SIM stack is accomplished in less than 200 ms on a modern graphics processing unit, enabling future applications for real-time imaging. Source code and ready-to-use software for the method are available at <http://ML-SIM.github.io>.

Published by The Optical Society under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

1. Introduction

Structured illumination microscopy (SIM) is an optical super-resolution imaging technique that was proposed more than a decade ago [1–5], and continues to stand as a powerful alternative to techniques such as Single Molecule Localization Microscopy (SMLM) [6,7] and Stimulated Emission Depletion (STED) microscopy [8]. The principle of SIM is that by illuminating a fluorescent sample with a patterned illumination, interference patterns are generated that contain information about the fine details of the sample structure that are unobservable in diffraction-limited imaging. In the simplest case of a sinusoidal illumination pattern with a spatial frequency of k_0 , the images acquired are a superposition of three copies of the sample's

frequency spectrum, shifted by $+k_0$, 0, and $-k_0$. The super-resolution image is reconstructed by isolating the three superimposed spectra and shifting them into their correct location in frequency space. The resulting spectrum is then transformed back into real space, leading to an image that is doubled in resolution. Isolating the three frequency spectra is mathematically analogous to solving three simultaneous equations. This requires the acquisition of three raw images, with the phase of the SIM patterns shifted with respect to one another along the direction of k_0 . Ideally, these phase shifts are in increments of $2\pi/3$ to ensure that the averaged illumination, i.e. the sum of all patterns, yields a homogeneous illumination field. Finally, to obtain isotropic resolution enhancement in all directions, this process is repeated twice, rotating the patterns by $2\pi/3$ each time, to yield a total of 9 images (i.e. 3 phase shifts for each of the 3 pattern orientations).

While SIM can be extended to resolve features down to the 50-60 nm range [9,10], it does not offer the highest resolution of the available super-resolution methods. However, the inherent speed of SIM makes it uniquely suited for live-cell imaging [11,12]. SIM also requires relatively low illumination intensities, and therefore reduces photo-toxicity and photo-bleaching compared to other methods. Many of the drawbacks of SIM relate to the reconstruction process, which can be time-consuming and prone to artefacts. In all but optimal imaging conditions, deviations from the expected imaging model or the incorrect estimation of experimental parameters (pixel size, wavelength, optical transfer function, image filters, phase step size etc.) introduce artefacts, degrading the final image quality [13]. This becomes especially prominent for images with low signal-to-noise ratios, where traditional methods will mistakenly reconstruct noise as signal, leading to artefacts that can be hard to distinguish from real features in the sample. At worst, the reconstruction process fails completely. These issues can introduce an element of subjectivity into the reconstruction process, leading to a temptation to adjust reconstruction parameters until the 'expected' result is obtained. In addition, traditional reconstruction methods are computationally demanding. The processing time for a single reconstruction in popular implementations such as FairSIM [14], a plugin for ImageJ/Fiji, and OpenSIM running in MATLAB [15], can reach tens of seconds even on high-end machines, making real-time processing during SIM image acquisition infeasible. Finally, traditional methods cannot easily reconstruct images from SIM data that is underdetermined, e.g. inputs with fewer than 9 frames and / or recordings with uneven phase steps between frames. These drawbacks limit the applicability of SIM when imaging highly dynamic processes [16]. Examples include the peristaltic movement of the endoplasmic reticulum [17] or the process of cell division [18], which require low light level imaging at high speed to reduce the effects of photo-toxicity and photo-bleaching.

In this work, we propose a versatile reconstruction method, ML-SIM, that addresses these issues with transfer learning. Transfer learning is a branch of machine learning that aims to exploit the knowledge obtained in an auxiliary domain to facilitate solving a specific task in the target domain [19]. While there are several methods to achieve this, a modern approach is to train a deep neural network to solve a similar task on a large dataset in the auxiliary domain, after which the network can be fine-tuned by slight changes to network architecture and retrained on a much smaller set of examples from the target domain [20]. ML-SIM uses an end-to-end deep residual neural network that is trained in an auxiliary domain consisting of simulated images using a high degree of randomisation. The training in the auxiliary domain is, in our case, sufficient for the network to generalise to a wide range of practically encountered conditions. This means that further fine-tuning of the model by training on real-world datasets, e.g. obtained from actual SIM experiments, is mostly not necessary. However, further fine-tuning and retraining is possible and supported by ML-SIM, thus offering maximal flexibility of the code to work for any experimental SIM implementation. Importantly, no output images from traditional reconstruction methods are required for training, thereby avoiding having the network undesirably learn to reproduce the reconstruction artefacts affecting the traditional methods. In a recent study [21], the problem of performing SIM reconstruction with a neural network, using U-Net

[22], was attempted in exactly this manner of using traditional reconstructed outputs as targets for training, thus simply approximating the current methods and prohibiting the network from becoming superior. Furthermore, the proposed deep residual network of ML-SIM is found to be significantly more capable of SIM reconstruction than the simpler U-Net – see Section 3.3. The training data in the auxiliary domain is generated by a simulation of the SIM imaging process. It is due to these training pairs of synthetic inputs and ideal, high-resolution targets (ground truths) for supervised learning that ML-SIM avoids exposing the model to the traditional reconstruction artefacts. Although the training data used is simulated and unrelated to real microscopic samples, we find that the method indeed generalises beyond the auxiliary domain, and we demonstrate successful application to experimental data obtained from two distinct SIM microscopes. This greatly empowers the method in the context of generalised reconstruction for super-resolution SIM imaging, since models can be customised to SIM setups of any configuration by changing simulation parameters used in the generation of the training data.

2. Methods

2.1. Convolution neural networks

Artificial neural networks consist of a sequence of layers that each performs a simple operation, typically a weighted sum followed by a nonlinear activation function, where every weight corresponds to a neuron in the layer. The weights are trainable, meaning that they are updated after every evaluation of an input performed during training. The updating scheme can be as simple as gradient descent with gradients determined via backpropagation of a loss calculated as the deviation between the network's output and a known target. A convolutional layer is no different, but utilises spatial information by only applying filters to patches of neighbouring pixels. The number of learned filters in one layer is a parameter, but is typically a power of 2, such as 32, 64 or 128. The network links past layers to present layers by skip connections to avoid the vanishing gradient problem. This type of architecture is known as a residual neural network [23].

Motivated by the results summarised in Fig. 7, and with the certainty that the entire input stack is utilised for the output reconstruction, the RCAN architecture was chosen for ML-SIM. The depth of the network was chosen to be around 100 convolutional layers (10 residual groups with 3 residual blocks). The network was then trained for 200 epochs with a learning rate of 10^{-4} , which was halved after every 20 epochs, using the Adam optimiser [24]. The models were implemented with Pytorch and trained using an Nvidia Tesla K80 GPU for approximately a day per model. Models have been trained on the full DIV2K image dataset except for a small collection of randomly selected images used for validation during training.

2.2. Generating simulated data

The ideal optical transfer function is generated based on a given objective numerical aperture (NA), pixel size and fluorescence emission wavelength. The illumination stripe patterns were then calculated from their spatial frequency k_0 and a phase ϕ ,

$$I_{\theta,\phi}(x, y) = I_0 \left[1 - \frac{m}{2} \cos(2\pi(k_x \cdot x + k_y \cdot y) + \phi) \right], \quad (1)$$

where $[k_x, k_y] = [k_0 \cos \theta, k_0 \sin \theta]$ for a pattern orientation θ relative to the horizontal axis. ϕ defines the phase of the pattern (i.e. the lateral shift in the direction of k_0), and m is the modulation depth which defines the relative strength of the super-resolution information contained in the raw images. In total, 9 images were generated for each target image, corresponding to three phase shifts for each of three pattern orientations – see Figure S3 for a depiction. The use of ML-SIM with different configurations of phase shifts and pattern orientations is covered in Supplement 1. The fluorescent response of the sample can then be modelled by the multiplication of the sample structure, $S(x, y)$ (input image), and the illumination pattern intensity $I_{\theta,\phi}(x, y)$. The final image,

$D_{\theta,\phi}(x, y)$, is then blurred by the PSF, $H(x, y)$, and noised with the addition of white Gaussian noise, $N(x, y)$,

$$D_{\theta,\phi}(x, y) = [S(x, y)I_{\theta,\phi}] \otimes H(x, y) + N(x, y), \quad (2)$$

where \otimes is the convolution operation. The option of using Poisson noise rather than Gaussian noise is explored in [Supplement 1](#). In addition to the Gaussian noise, $N(x, y)$, added pixel-by-pixel, a random error is added to the parameters for the stripe patterns, k_0 , θ and ϕ , to approximate the inherent uncertainty in an experimental setup for illumination pattern generation. The importance of including these types of errors is described in [Supplement 1](#).

The images generated from 2 are used as inputs in a supervised learning approach. The targets used to calculate loss for optimising the neural network are the clean grayscale source images used as $S(x, y)$. These targets could as well be blurred with a PSF corresponding to the best theoretically achievable resolution of standard SIM, but as explored in [Supplement 1](#) this is not found to be beneficial and instead the unmodified source images are used as targets and referred to as the ground truths. The values for m and k_0 that are used for data generation are given in [Supplement 1](#).

2.3. Microscopy

For the experimental data described in Section 3, two custom-built SIM microscopes were used. For the imaging of the endoplasmic reticulum (ER), a SIM instrument based on a phase-only spatial light modulator was used. The microscope objective used was a 60X 1.2 NA water immersion lens and fluorescence was imaged with a sCMOS camera. Cells were labelled with VAPA-GFP and excited by 488 nm laser light. For the imaging of the cell membrane, a novel SIM setup based on interferometry for the pattern generation was used [25]. In this system, the angle and phase shifts are achieved by rotation of a scanning mirror, the repeatability of which introduces uncertainty into the phase shifting. The microscope objective used was a 60X 1.2 NA water immersion lens, and fluorescence was imaged with an sCMOS camera. The cell membrane was stained with a CAAX-Venus label and excited with 491 nm laser light. On both systems, 200 nm beads labelled with Rhodamine B were excited by 561 nm laser light. For both images, the traditional reconstruction methods that have been tested managed to reconstruct the raw SIM stacks, although with varying success for the interferometric SIM setup due to the irregularity of the phase stepping.

3. Results

3.1. Using machine learning to train a reconstruction model

ML-SIM is built on an artificial neural network. Its purpose is to process a sequence of raw SIM frames (i.e. a stack of nine images representing the number of raw images acquired in a SIM experiment), into a single super-resolved image. To achieve this, a supervised learning approach is used to train the network, where pairs of inputs and desired outputs (ground truths) are presented during training to learn a mapping function. These training data could be acquired on a real SIM system using a diverse collection of samples and experimental conditions. However, the targets corresponding to the required inputs are more difficult to obtain. At least two approaches seem possible: (A) using outputs from traditional reconstruction methods as targets [21]; and (B) using images from other super-resolution microscopy techniques that are able to achieve higher resolution than SIM (e.g SMLM or STED). Option (A) would prohibit ML-SIM from producing reconstructions that surpass the quality of traditional methods and would be prone to reproducing the artefacts mentioned in Section 1. Option (B) requires a capability to perform correlative imaging of the same sample, which may be difficult to achieve since training requires hundreds or even thousands of distinct data pairs [29]. In addition, both approaches require the preparation of many unique samples to build a training set diverse enough for the model to generalise well.

Hence, these options were not pursued in this work and we approached the problem instead by starting with ground truth images, and simulating inputs by mimicking the SIM process in silico, allowing for very diverse training sets to be built. We used the image set DIV2K [30], which consists of 1000 high-resolution images of a large variety of objects, patterns and environments. To generate the SIM data, images from the image set were first resized to a standard resolution of 512x512 pixels and transformed to greyscale. Raw SIM images were then calculated using a SIM model adapted from the OpenSIM package [15]. The model and underlying parameters are described in Section 2. The simulated raw SIM stacks were used as input to the neural network and the output compared to the known ground truth in order to calculate a loss to update the network weights. Figure 1 shows an overview of the training process with an example of a simulated SIM input. The architecture of the neural network is further described in Supplement 1.

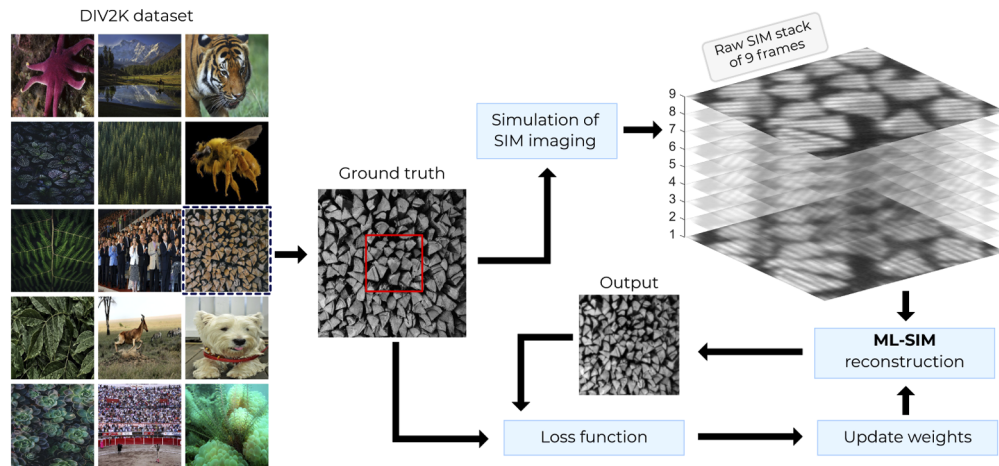


Fig. 1. Data processing pipeline for ML-SIM. Training data for the model is generated by taking images from the commonly used image dataset DIV2K and simulating the imaging process of SIM using a model adapted from the open source library OpenSIM. The simulation can be further optimised to reflect the properties of the experimental system for which the reconstruction method is desired, for example to match the pixel size of the detector or numerical aperture of the detection optics. The outputs of the simulation are image stacks of the same size as those acquired by the microscope (here 9 frames).

3.2. Application of the trained model

To begin with, we tested that the network had learned to reconstruct simulated SIM stacks. Prior to training, a separate partition of DIV2K was selected for testing. A sample from this test partition is shown in Fig. 2. The stripe pattern for two of the nine frames of the input SIM stack are shown in the leftmost panel. The stripe patterns cancel out when all 9 frames are summed together (second column), and this corresponds to the case of even illumination in a wide-field microscope. Compared to the wide-field image, the reconstruction from ML-SIM is seen to have a much improved resolution with a peak signal-to-noise ratio (PSNR) value more than 7 dB higher as well as a significantly higher structural similarity index (SSIM). Beyond these metrics, several features of the image can be seen to be resolved after reconstruction that were not visible beforehand, such as the vertical posts seen to the right side of the cropped region.

It should also be noted that the reconstruction has not amplified the noise by introducing any evident artefacts, even though the input image featured a significant amount of Gaussian noise in addition to randomisation of the stripe frequency and phase – see Section 2 for definitions of

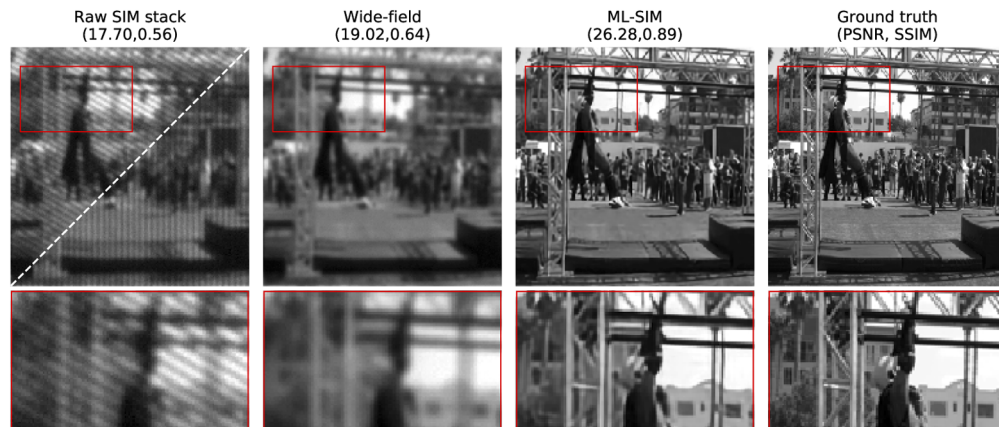


Fig. 2. Generation of training datasets for ML-SIM. Column 1: Sample from test partition of DIV2K (ground truth) transformed to a raw data stack of 9 frames via simulation of the SIM imaging process. Two different orientations are shown for the excitation patterns. Column 2: Wide-field image, obtained as the mean of the 9 raw frames. Column 3: Super-resolved image obtained through reconstruction with ML-SIM. Column 4: Ground truth. The image quality metrics shown in brackets are the peak signal-to-noise ratio and the structural similarity index [26], respectively.

those parameters. As further described in Section 2, the neural network underlying ML-SIM is different to those of generative networks, which means that the model is more strongly penalised during training for introducing image content that is not in the real image. We argue that, even though this results in slightly more blurred output images than would be achievable with a generative network [31,32], the absence of artificial features is preferable in scientific imaging applications. This trade-off is referred to as minimising the mathematical reconstruction error (e.g. root-mean-square deviation) rather than optimising the perceptual quality [33,34].

While ML-SIM is able to reconstruct simulated SIM stack inputs, it is of course only valuable if it also works on real SIM data, acquired experimentally. The ML-SIM model was trained on input data from simulations, using data bearing little resemblance to real-world biological SIM data. Any success for real-world SIM reconstructions therefore requires the model to have generalised the SIM process in such a way that it becomes independent of image content and sample type. This requires a realistic simulation of the SIM imaging process to generate training data that is sufficiently diverse on the one hand, and reflects measurement imperfections as encountered in practical SIM imaging. The former was avoided through use of a diverse training dataset, and the latter through use of the well-known imaging response function (Section 2, Eq. (2)), and introduction of uncertainty in the stripe patterns. To test ML-SIM on experimental data, SIM images of different samples were acquired with two different SIM setups [35]. The resulting reconstructed outputs are shown in Fig. 3, where they are compared to outputs of traditional reconstruction methods: OpenSIM [15], a cross-correlation (CC-SIM) phase retrieval approach [28,36], and FairSIM [14]. The images are grayscale images of signal intensity mapped to the Viridis colour table. ML-SIM is seen to obtain resolution on par with the other methods but producing less noisy background and fewer artefacts. The bottom two rows of images of beads and cell membranes were acquired with phase steps deviating from the ideal $2\pi/3$. This reflects a difficulty with the interferometric SIM setup (see Section 2) to achieve equidistant, and precisely defined, phase steps for each illumination pattern angle. This means that the reconstruction algorithm must handle inconsistent phase changes, a factor only the cross-correlation method was capable of handling. However, although CC-SIM has improved resolution, artefacts are apparent,

seen as vertical lines and ringing in the images. ML-SIM, on the other hand, reconstructed with fewer artefacts and strongly improved background rejection.

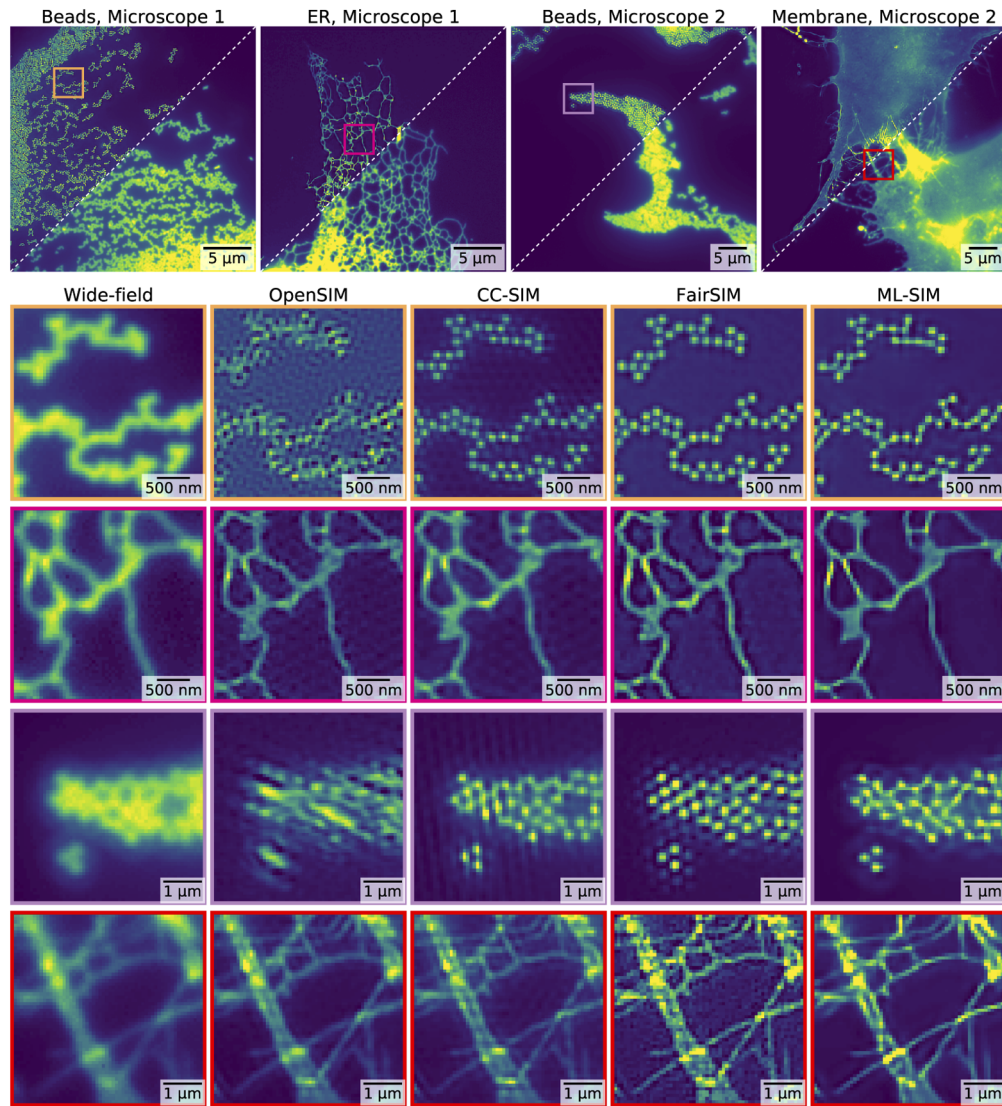


Fig. 3. Reconstruction of SIM images from four different samples imaged on two different experimental SIM set-ups. Microscope 1 uses a spatial light modulator for stripe pattern generation [27], while microscope 2 uses interferometric pattern generation. Both instruments were used to image a sample consisting of fluorescent beads as well as biological samples featuring the endoplasmic reticulum (ER) and a cell membrane, respectively. (Top) Full field-of-view images where each upper left half shows the reconstruction output from ML-SIM and each lower right half shows the wide-field version taken as the mean of the raw SIM stack. (Bottom) Cropped regions of reconstruction outputs from OpenSIM [15], CC-SIM [28], FairSIM [14] and ML-SIM. Panels in rows 2 to 5 correspond to regions indicated by coloured boxes in the full frame images.

To further demonstrate the super-resolution performance of ML-SIM, a sample of 30 nm microtubules labelled with Alexa-647 was imaged on microscope 1. The reconstruction outputs

and line profiles across neighbouring microtubules for both ML-SIM and FairSIM are shown on Fig. 4. The displayed cropped region contains two parallel microtubules which are separated by a gap of size below the diffraction limit and thus not resolved in the wide-field image. In the outputs from ML-SIM and FairSIM the gap is clearly visible. The distance between the peaks in the line profile for ML-SIM and FairSIM is ≈ 150 nm, which is close to the theoretically achievable resolution with standard SIM [3]. Analysis of the resulting OTFs after reconstruction is also provided in [Supplement 1](#).

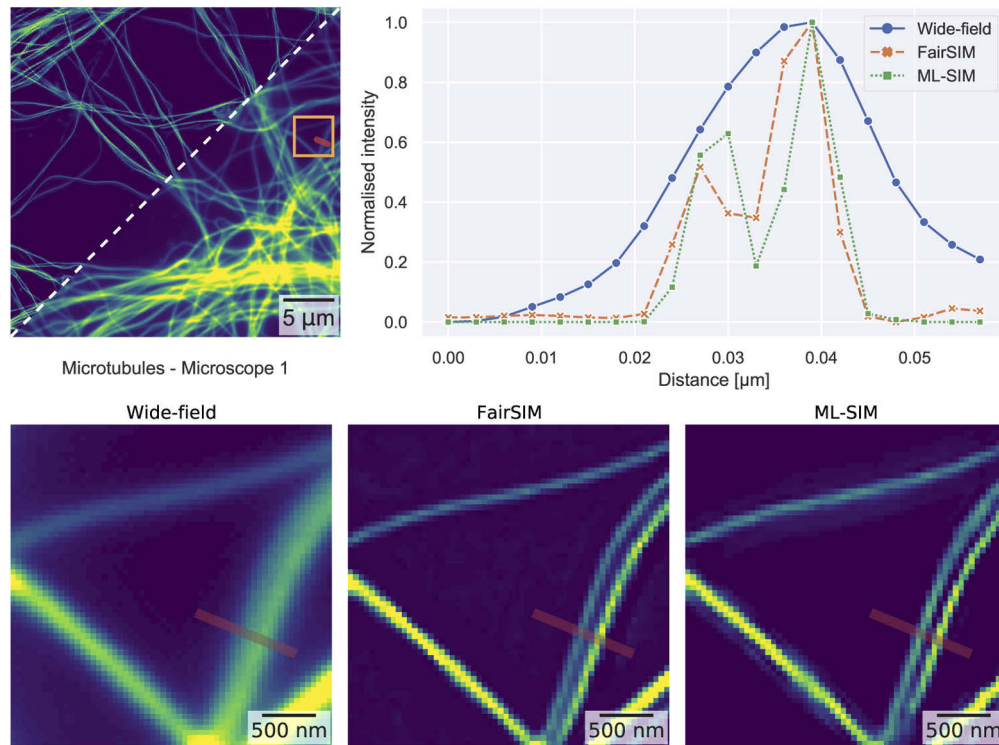


Fig. 4. Reconstruction of a SIM image of tubulin structures. The reconstruction output of ML-SIM is compared with a wide-field projected image and FairSIM. (Top) Full field-of-view of reconstructed image and line profiles across two parallel microtubules at the position indicated by the red line. While the microtubules are not resolved in widefield mode, both ML-SIM and FairSIM enable them to be clearly distinguished. (Bottom) Cropped regions of the reconstruction outputs corresponding to the area enclosed by the yellow rectangle.

The application of ML-SIM to TIRF-SIM image data using a sample image from the official FairSIM test image repository is described in [Supplement 1](#).

3.3. Performance assessment

We performed a quantitative comparison of ML-SIM with traditional reconstruction methods on reconstructions of simulated raw SIM stacks generated from two image datasets; a subset of 10 DIV2K images, unseen during training, and 24 images from a dataset referred to as Kodak 24, commonly used for image restoration benchmarking [30,37]. Parameters for OpenSIM, CC-SIM and FairSIM were all systematically adjusted to produce the highest achievable output quality. Consequently, each method required completely different parameter configurations than those used for reconstructions of the experimental data shown in Fig. 3. For ML-SIM however, there were no tunable parameters. The optical transfer function (OTF) is estimated within each method

even though the function is known for the simulated images – this is the same premise as for the reconstruction of the experimental samples in Fig. 3, for which the OTFs were unknown. Each method applies an identical Wiener filter to the final reconstruction output, whereas the output of ML-SIM is untouched. The performance scores for all methods measured in PSNR and SSIM averaged over the entire image sets are listed in Table S1 with scores for wide-field as a reference in terms of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). For both metrics, ML-SIM has the highest scores with a PSNR that is 2 dB higher than for OpenSIM. CC-SIM and FairSIM lag behind but both methods still succeed in improving the input beyond the baseline wide-field reference. The performance gap between OpenSIM and the other traditional methods is likely due to a better estimation of the OTF, because OpenSIM assumes an OTF that is similar to the one used when simulating the SIM data.

A more challenging test image than those based on DIV2K and Kodak 24 images is shown in Fig. 5. This simulated test image is reconstructed with the same three traditional methods. OpenSIM is found to achieve the best reconstruction quality of the three with a PSNR score of 13.84 dB versus 12.56 dB for CC-SIM and 12.88 dB for FairSIM. The same image reconstructed with ML-SIM results in a PSNR score of 16.32 dB – again about 2 dB higher than OpenSIM. Two cropped regions comparing OpenSIM and ML-SIM are shown in Fig. 5. The area in the upper right corner of the test image is particularly challenging to recover due to the single-pixel point patterns and the densely spaced vertical lines. While the points vanish in the wide-field image, these are recovered both by OpenSIM and ML-SIM. The resolution of the point sources are slightly superior in the ML-SIM reconstruction, and ML-SIM manages to recover the high-frequency information in the top line pattern very well. Overall it is also seen that the reconstruction from ML-SIM contains much less noise, which is especially evident in the zoomed region of the face. This suggests that ML-SIM is less prone to amplify noise present in the input image. We tested this further by gradually adding more Gaussian image noise to the input image, and again comparing the reconstructions from the various methods. The results of this test are shown in Fig. 6, where it is seen clearly that ML-SIM performs best at high noise levels. As more noise is added the gap in performance is seen to increase between ML-SIM and the other models indicating that the neural network has learned to perform denoising as part of the reconstruction process. This is supported by the cropped regions on the right side of the figure, which show cleaner detail in the image when compared to the input, wide-field and OpenSIM images. OpenSIM was found to perform consistently well in this noise test, whereas FairSIM and CC-SIM struggled to reconstruct at all for higher noise levels. This is not surprising, since added noise may cause the parameter estimation to converge to incorrect optima, which can heavily corrupt the reconstruction outputs. As a result, the reconstruction outputs from FairSIM and CC-SIM were of worse quality than the wide-field reference at higher noise levels.

Several architectures were tested as part of this research to select the one most suitable for ML-SIM. U-Net [22] is a popular, versatile and easily trained network, but its performance was found to fall short of state-of-the-art single image super-resolution networks such as EDSR [38] and RCAN [39]. These super-resolution networks have been customised to be able to handle input stacks of up to 9 frames and output a single frame with no upsampling, i.e. the upsampling modules of those networks have been omitted – see Figure S2 for a depiction. In addition to testing different network architectures the number of frames of the input raw SIM stack, up to a total of 9, was also varied. In the left-hand side of Fig. 7 the convergence of test scores on a validation set during training are shown for the various architectures and input configurations considered. It is found that SIM reconstruction with subsets containing only 3 or 6 frames still performed significantly better than if the network learns to perform a more simple deconvolution operation by just training on a wide-field input. This confirms that the network learns to extract information from all 9 frames in the full stack versus a subset of it or the mean of its frames. Only using a subset of 3 frames does however cause a substantial reconstruction quality loss compared

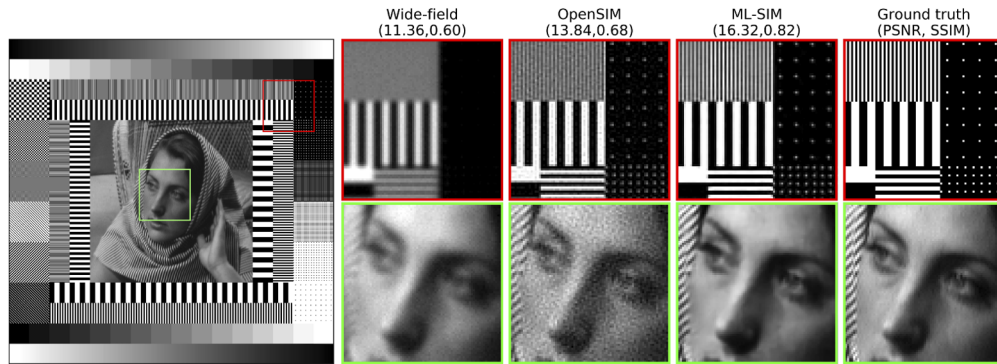


Fig. 5. Reconstructions of a test target with OpenSIM and ML-SIM and comparison to the ground truth. OpenSIM was found to be the best performing traditional method on this test sample, both in terms of PSNR and SSIM with the other methods achieving PSNR scores of 12.56 dB (CC-SIM) and 12.88 dB (FairSIM).

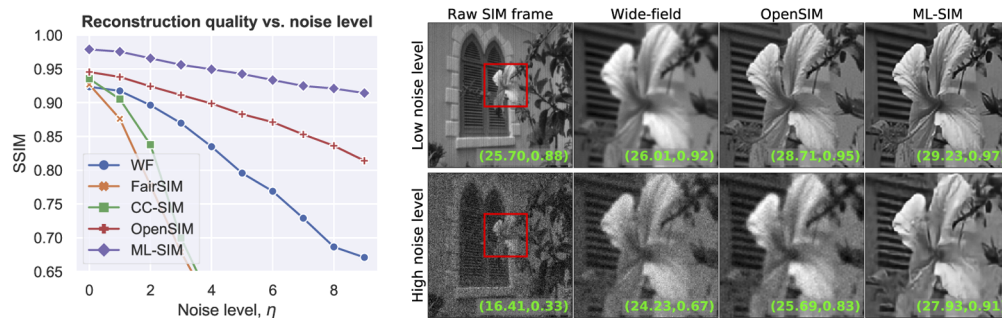


Fig. 6. (Left) Reconstruction quality as measured by the structural similarity index, SSIM, as a function of the amount of noise added to an input image. Gaussian noise is added to every frame of the raw SIM stack. Noise is normally distributed with a standard deviation $\eta \cdot \sigma$, where σ is the standard deviation of the input image. (Right) Images at low ($\eta = 0$) and high noise levels ($\eta = 9$) reconstructed with OpenSIM and ML-SIM, respectively. PSNR and SSIM scores using the ground truth as reference are shown in the lower right hand corner of every image.

to using 6 frames, which is not surprising since the corresponding analytical reconstruction problem becomes underdetermined for fewer than 4 frames [16]. The RCAN model performs better than EDSR with a consistently higher PSNR score when trained on all 9 frames, while performing similarly to EDSR when trained with 3 fewer frames. Based on these results RCAN is chosen as the default architecture for ML-SIM. The fact that reconstruction with fewer than 9 frames is possible could be exploited for compressed, faster SIM imaging as done in [21] using a U-Net model, although this inevitably comes at a loss of quality.

Regarding the computation time for each reconstruction method, we measured the average time for reconstructing the raw SIM stacks based on the Kodak 24 image dataset one by one. The timing for each method is then the mean of 24 time samples with an associated standard deviation. The timings are shown on the right-hand side of Fig. 7. Computations were performed on a computer running Windows 10 with an Intel i5 6500 CPU, 16 GB DDR4 RAM and a Nvidia GTX 1080 Ti GPU. ML-SIM finishes a reconstruction in less than 200 ms, which is more than an order of magnitude faster than the other methods. Substantial speedups are to be expected when using neural networks due to the computations being greatly parallelisable, thus making it

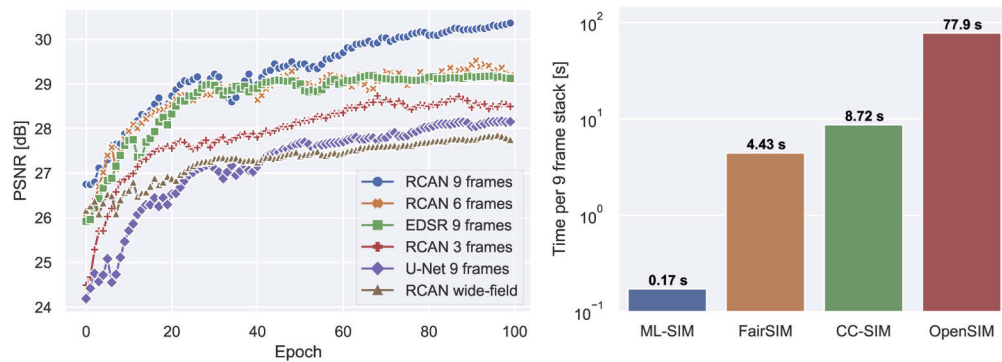


Fig. 7. (Left) Validation test set scores during training for different network architectures and input dimensions. The two state-of-the-art single image super-resolution architectures, RCAN and EDSR, have been modified to perform SIM reconstruction. The number of frames of the raw SIM stack, up to a total of 9, is also varied to confirm that the network learns to extract information from all 9 frames in the full stack. (Right) Computation time for reconstruction of a single raw SIM stack of 9 frames. The shown times are averages of 24 consecutive reconstructions with sample standard deviations of 0.0034, 0.13, 0.51 and 3.7 seconds for ML-SIM, FairSIM, CC-SIM and OpenSIM, respectively.

easy to use GPU acceleration – this was similarly found in [40], where a neural network was used for reconstruction of stochastic optical reconstruction microscopy images. The traditional methods for SIM reconstruction are more difficult to parallelise, partly because the numerical optimisation algorithms needed for parameter estimation tend to be iterative and sequential. This therefore provides a computational advantage of ML-SIM. At 170 ms per reconstructed image from a SIM stack of 9 frames, the reconstructed image rate is about 6 per second, corresponding to an imaging system that captures 54 frames per second, which could provide fluent, real-time, super-resolution feedback to the user during image acquisition.

3.4. Web app, desktop app and source code

The source code for training ML-SIM and applying the model for reconstruction is available in a public repository on GitHub, <https://github.com/charlesnchr/ML-SIM>, and figshare (Code 1) [41]. This repository includes source code for generating the training data by simulating the SIM imaging process with parameters that can be easily adapted to reflect specific SIM setups (e.g. by changing stripe orientations, number of frames, etc.). The repository also holds source code for a desktop program with pre-built installers for Windows, macOS and Linux. The program makes it easy to use ML-SIM and perform batch processing via a graphical user interface. During installation required dependencies such as Python, Pytorch and pre-trained ML-SIM models are automatically fetched. If the pre-trained models perform suboptimally, it is easy to train a new model that is more specific to a given SIM setup and set the program to use this custom model for reconstruction. The program includes a plugin for μ Manager [42] that enables a real-time live-view of ML-SIM reconstructed output during acquisition in many imaging systems thanks to the wide support of camera drivers in μ Manager. See Supplement 1 for more details. Furthermore, we have created a web app accessible via <http://ML-SIM.github.io> with a browser-based online implementation of ML-SIM that is ready for quick testing using a pre-trained model and does not require installation of any software.

4. Discussion

We demonstrate and validate a SIM reconstruction method, ML-SIM, which takes advantage of transfer learning by training a model in an auxiliary domain consisting of simulated images and generalises to the target task of reconstructing experimental SIM images with no fine-tuning or retraining necessary. The training data was generated by simulating raw SIM image data from images obtained from common image repositories, serving as ground truths. ML-SIM successfully reconstructed artificial test targets that were of a completely different nature than the diverse images used to generate the training datasets. More importantly, it successfully reconstructed real data obtained by two distinct experimental SIM implementations. We compared the performance of ML-SIM to widely used reconstruction methods, OpenSIM [15], FairSIM [14], and CC-SIM [28]. In all cases, reconstruction outputs from ML-SIM contained less noise and fewer artefacts, while achieving similar resolution improvements. Through a randomisation of phase shifts in the simulated training data, it was also possible to successfully reconstruct images that could not be processed successfully with two of the traditional reconstruction methods. ML-SIM shows robustness to unpredictable variations in the SIM imaging parameters and deviations from equidistant phase shifts. Similarly, ML-SIM reconstructed images that were strongly degraded by noise even beyond the point where the other methods failed.

A central advantage of the transfer learning approach of the ML-SIM method is that the simulated data that constitute the auxiliary domain can be made arbitrarily diverse by randomisation of optical parameters, enabling the model to become highly generalised for the target task. Furthermore, the simulation can also be optimised to a specific system by changing relevant optical parameters. In principle, this makes the method applicable to any SIM setup regardless of its configuration. For instance, a SIM setup with another illumination pattern configuration, e.g. 5 orientations and 5 phase shifts (5x5 stacks), is trivial to support with ML-SIM by changing just two parameters in the pipeline. General, pretrained models for SIM microscopes with configurations for 3x3, 3x5 and 5x5 stacks are provided at <http://ML-SIM.github.io> along with source code and software to use them.

A future direction could be to fine-tune the training data by incorporating a more sophisticated image formation model. This image formation model might also take certain optical aberrations into account. Currently, out-of-focus light from above and below the focal plane is not simulated in the training data. As with the other reconstruction methods, this can result in artefacts in regions of the sample with dense out-of-focus structures. Given that the spatial frequency information required to remove this background is available in SIM, it is possible that an updated ML-SIM network could be constructed that incorporates an efficient means for background rejection [43,44].

Funding. Engineering and Physical Sciences Research Council (L015889); Wellcome Trust (089703); Medical Research Council (K015850).

Acknowledgments. The authors thank Jerome Boulanger from the Laboratory of Molecular Biology in Cambridge, UK, for general guidance on reconstruction for SIM. The authors also thank group member Lisa Hecker from the Laser Analytics Group of the University of Cambridge for providing experimental data from the SLM and interferometric SIM microscopes described in this study.

Disclosures. The authors declare no conflicts of interest.

Supplemental document. See [Supplement 1](#) for supporting content.

References

1. C. J. R. Sheppard, "Super-resolution in confocal imaging," *Optik* **80**, 53–54 (1988).
2. R. Heintzmann and C. G. Cremer, "Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating," in *Optical Biopsies and Microscopic Techniques III*, vol. 3568 (SPIE, 1999), pp. 185–196.
3. M. G. Gustafsson, "Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy," *J. Microsc.* **198**(2), 82–87 (2000).

4. M. G. L. Gustafsson, L. Shao, P. M. Carlton, C. J. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat, "Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination," *Biophys. J.* **94**(12), 4957–4970 (2008).
5. L. Schermelleh, R. Heintzmann, and H. Leonhardt, "A guide to super-resolution fluorescence microscopy," *J. Cell Biol.* **190**(2), 165–175 (2010).
6. W. E. Moerner and L. Kador, "Optical detection and spectroscopy of single molecules in a solid," *Phys. Rev. Lett.* **62**(21), 2535–2538 (1989).
7. E. Betzig, "Proposed method for molecular optical imaging," *Opt. Lett.* **20**(3), 237 (1995).
8. S. W. Hell and J. Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy," *Opt. Lett.* **19**(11), 780 (1994).
9. D. Li, L. Shao, B.-C. Chen, X. Zhang, M. Zhang, B. Moses, D. E. Milkie, J. R. Beach, J. A. Hammer, and M. Pasham, "Extended-resolution structured illumination imaging of endocytic and cytoskeletal dynamics," *Science* **349**(6251), aab3500 (2015).
10. E. H. Rego, L. Shao, J. J. Macklin, L. Winoto, G. A. Johansson, N. Kamps-Hughes, M. W. Davidson, and M. G. Gustafsson, "Nonlinear structured-illumination microscopy with a photoswitchable protein reveals cellular structures at 50-nm resolution," *Proc. Natl. Acad. Sci.* **109**(3), E135–E143 (2012).
11. F. Ströhl and C. F. Kaminski, "Frontiers in structured illumination microscopy," *Optica* **3**(6), 667 (2016).
12. P. W. Winter and H. Shroff, "Faster fluorescence microscopy: advances in high speed biological imaging," *Curr. Opin. Chem. Biol.* **20**, 46–53 (2014).
13. G. Ball, J. Demmerle, R. Kaufmann, I. Davis, I. M. Dobbie, and L. Schermelleh, "SIMcheck: a toolbox for successful super-resolution structured illumination microscopy," *Sci. Rep.* **5**(1), 15915 (2015).
14. M. Müller, V. Mönkemöller, S. Hennig, W. Hübner, and T. Huser, "Open-source image reconstruction of super-resolution structured illumination microscopy data in ImageJ," *Nat. Commun.* **7**(1), 10980 (2016).
15. A. Lal, C. Shan, and P. Xi, "Structured illumination microscopy image reconstruction algorithm," *IEEE J. Select. Topics Quantum Electron.* **22**(4), 50–63 (2016).
16. F. Ströhl and C. F. Kaminski, "Speed limits of structured illumination microscopy," *Opt. Lett.* **42**(13), 2511 (2017).
17. D. Holcman, P. Parutto, J. E. Chambers, M. Fantham, L. J. Young, S. J. Marciniak, C. F. Kaminski, D. Ron, and E. Avezov, "Single particle trajectories reveal active endoplasmic reticulum luminal flow," *Nat. Cell Biol.* **20**(10), 1118–1125 (2018).
18. T. A. Planchon, L. Gao, D. E. Milkie, M. W. Davidson, J. A. Galbraith, C. G. Galbraith, and E. Betzig, "Rapid three-dimensional isotropic imaging of living cells using Bessel beam plane illumination," *Nat. Methods* **8**(5), 417–423 (2011).
19. S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2010).
20. J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowledge-Based Systems* **80**, 14–23 (2015).
21. L. Jin, B. Liu, F. Zhao, S. Hahn, B. Dong, R. Song, T. C. Elston, Y. Xu, and K. M. Hahn, "Deep learning enables structured illumination microscopy with low light levels and enhanced speed," *Nat. Commun.* **11**(1), 1934 (2020).
22. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Springer, 2015), pp. 234–241.
23. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016-Decem.*, 770–778 (2016).
24. D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," Tech. rep. (2015).
25. W. Liu, Q. Liu, Z. Zhang, Y. Han, C. Kuang, L. Xu, H. Yang, and X. Liu, "Three-dimensional super-resolution imaging of live whole cells using galvanometer-based structured illumination microscopy," *Opt. Express* **27**(5), 7237–7248 (2019).
26. Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. on Image Process.* **13**(4), 600–612 (2004).
27. R. Fiolka, M. Beck, and A. Stemmer, "Structured illumination in total internal reflection fluorescence microscopy using a spatial light modulator," *Opt. Lett.* **33**(14), 1629–1631 (2008).
28. K. Wicker, O. Mandula, G. Best, R. Fiolka, and R. Heintzmann, "Phase optimisation for structured illumination microscopy," *Opt. Express* **21**(2), 2032 (2013).
29. M. Hauser, M. Wojcik, D. Kim, M. Mahmoudi, W. Li, and K. Xu, "Correlative super-resolution microscopy: new dimensions and new opportunities," *Chem. Rev.* **117**(11), 7428–7456 (2017).
30. E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: dataset and study," Tech. rep. (2017).
31. C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 105–114 (2017).
32. X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, "ESRGAN: enhanced super-resolution generative adversarial networks," Tech. rep. (2018).
33. Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "2018 PIRM challenge on perceptual image super-resolution," Tech. rep. (2018).

34. S. Vasu, N. T. Madam, and A. N. Rajagopalan, "Analyzing perception-distortion tradeoff using enhanced perceptual super-resolution network," Tech. rep. (2018).
35. L. J. Young, F. Ströhl, and C. F. Kaminski, "A guide to structured illumination TIRF microscopy at high speed with multiple colors," *J. Vis. Exp.* **111**, 53988 (2016).
36. R. Cao, Y. Chen, W. Liu, D. Zhu, C. Kuang, Y. Xu, and X. Liu, "Inverse matrix based phase estimation algorithm for structured illumination microscopy," *Biomed. Opt. Express* **9**(10), 5037 (2018).
37. J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," arXiv:1803.04189 (2018).
38. B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," Tech. rep. (2017).
39. Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," Tech. rep. (2018).
40. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," *Optica* **5**(4), 458 (2018).
41. C. N. Christensen, "ML-SIM," figshare (2021). <https://figshare.com/articles/software/ML-SIM/13702195>.
42. A. Edelstein, N. Amodaj, K. Hoover, R. Vale, and N. Stuurman, "Computer control of microscopes using μ manager," *Curr. Protoc. Mol. Biol.* **92**(1), 14–20 (2010).
43. J. Qian, M. Lei, D. Dan, B. Yao, X. Zhou, Y. Yang, S. Yan, J. Min, and X. Yu, "Full-color structured illumination optical sectioning microscopy," *Sci. Rep.* **5**(1), 14513 (2015).
44. M. Neil, R. Juškaitis, and T. Wilson, "Real time 3D fluorescence microscopy by two beam interference illumination," *Opt. Commun.* **153**(1-3), 1–4 (1998).